The Site-Specific Recombination System of the *Escherichia coli* Bacteriophage Φ24B

Mohaisen, M.R [1,2*], McCarthy, A.J.[1], Adriaenssens, E.M.[3] Allison, H.E.[1*]

[1]Department of Functional and Comparative Genomics, Institute of Integrative Biology, University of Liverpool, Liverpool, UK

[2]College of Dentistry, Anbar University, Anbar, Iraq
[3]Quadram Institute Bioscience, Norwich, Norfolk, UK

*** Correspondence:**
Heather E. Allison
hallison@liverpool.ac.uk;

Mohammed R. Mohaisen
den.m.radhi@uoanbar.edu.iq

Keywords: Shiga toxin, Site-Specific DNA Recombination, Stx phage, Prophage, Prophage integrase, tyrosine recombinase, *attP*, *attB*

1 Abstract

Stx bacteriophages are members of the lambdoid group of phages and are responsible for Shiga toxin (Stx) production and the dissemination of Shiga toxin genes (*stx*) across shigatoxigenic *E. coli* (STEC). These toxigenic bacteriophage hosts can cause life-threatening illness, and Shiga toxin (Stx) is the virulence determinant responsible for the severe nature of infection with enterohemorrhagic *E. coli* (EHEC), a subset of pathogenic STEC. Stx phages are temperate, and in the present study the identification of what is actually required for Stx phage Φ24B and bacterial DNA recombination was tested using both *in vitro* and *in situ* recombination assays. It is well established that phage λ, which underpins most of what we understand about lambdoid phage biology, requires its own encoded phage attachment site (*attP*) of 250 bp, a host encoded attachment site (*attB*) of 21 bp and a host encoded, DNA binding protein known as integration host factor (IHF). The assays applied in this study enabled the manipulation of the phage attachment site (*attP*) and bacterial attachment site (*attB*) sequences, and the inclusion or exclusion of a host-encoded accessory element known as integration host factor. We were able to demonstrate that the minimal *attP* sequence required by Φ24B phage is between 350 and 427 bp. Unlike phage λ, the minimal necessary flanking sequences for the *attB* site do not appear to be equal in size with a total length between 62 and 93 bp. Furthermore, we identified that the Φ24B integrase does not require IHF to drive the integration and recombination process. Understanding how this unusual Stx phage integrase works may enable exploitation of its promiscuous nature in the context of genetic engineering.

2 Introduction

Shigatoxigenic *Escherichia coli* (STEC) emerged in the early 1980s as the causal agents of a variety of clinical symptoms and sequelae ranging from mild diarrhea through to life threatening conditions such as hemolytic uremic syndrome (HUS) and thrombotic thrombocytopenic purpura (TTP)(Fogg et al, 2011; Paton & Paton, 1998). The key virulence factor of STEC is the expression of Shiga toxin (Stx), which is an AB_5 toxin encoded on a small operon comprising 2 genes. The *stx* genes are acquired following infection with an Stx phage (O'Brien et al, 1984), a temperate lambdoid phage that carry *stx* genes within the late gene region of its genome (Smith et al, 2012; Unkmeir & Schmidt, 2000). Stx phages are members of the lambdoid group of phages which have a common genomic organization and content. The behavior of the archetypal member of this group, phage λ, underpins most of what we understand about lambdoid phage biology. The expression of the *stx* genes is linked to the bacteriophage's lytic replication cycle, which subverts the host cell's resources in order to produce more phage particles (Allison, 2007; Unkmeir & Schmidt, 2000; Wagner et al, 2002). However, not all bacteriophage infections result in the lysis of the host cell. DNA from most , temperate phages, including phage λ and Stx phages, can, alternatively, become integrated in the host genome, becoming a prophage (Madsen et al, 1999; Saunders et al, 2001) or in a few other phages survive in the cell as a replicating plasmid (Łobocka et al, 2004). For those phage destined to become integrated, the incoming phage produces a site-specific recombinase enzyme that directs phage recombination within the bacterial genome. For lambdoid phages, this recombinase enzyme is known as integrase (Groth & Calos, 2004).

The integrase from phage λ (Intλ) is a tyrosine, site-specific recombinase enzyme that drives recombination between two specific, complimentary DNA sequences, the *attPλ* (250 bp) site (Biswas et al, 2005; Hsu et al, 1980; Mizuuchi & Mizuuchi, 1980) located on the phage genome and the *attBλ* site (21 bp) (Mizuuchi & Mizuuchi, 1985) located within the bacterial genome (Fogg et al, 2014; Mizuuchi & Mizuuchi, 1980). The *attPλ* is composed of two integrase binding sites (P and P') and the core-binding site (COC'). The *attBλ* site comprises a central overlap region, known also as O (7 bp) which is flanked by imperfect, inverted repeats, B (7 bp) and B'(7 bp) (Mizuuchi & Mizuuchi, 1985). Though Intλ is the only phage-encoded protein needed for phage λ integration within the host genome (Fogg et al, 2011), most tyrosine integrases do not act autonomously and require the help of a bacterial encoded accessory factor (Groth & Calos, 2004). The host-encoded accessory factor required by Intλ is integration host factor IHF, which works as an accessory protein, enhancing the enzymatic activity Intλ. IHF, a ~ 21.8 kDa heterodimeric DNA-binding protein, is composed of two monomers: α (~11 kDa) and β (~9.5 kDa) (Azam & Ishihama, 1999; Kikuchi & Nash, 1978; Sanyal et al, 2014), encoded by *himA* and *himD* genes, respectively (Yang & Nash, 1995). IHF interacts with three specific DNA sequences surrounding the *attPλ* site in order to maintain a DNA topography that enables the action of the now active homotetrameric Intλ enzyme (Sugimura & Crothers, 2006). IHF binding sites are around 30 - 35 bp in size and are composed of at least two domains: the 3' region and the 5' region. Unlike the 3' region, the 5' region is random in most cases. Alignment among IHF binding sequences within an *attL* population library showed that the 3' region is where IHF binds strongly and specifically due to the presence of a 13 bp consensus sequence (WATCAANNNNTTR). The 5' sequences are all simply AT rich and lack any obvious sequence patterns (Goodman et al, 1999).

The crossover regions of *attBλ* and *attPλ* interact with each other through sequence homology (Weisberg et al, 1983) aided by protein-DNA interactions between IHF and specific sites in *attP* that lie between COC and P (Frumerie et al, 2008). The Intλ, *attBλ*, *attPλ*, and IHF complex then allows Intλ to sequentially exchange crossover regions of the *attB* and *attP* sites. Intλ possesses a high binding affinity for the P and P' sites of *attP* and low affinity binding to the COC' site of *attP* (Groth & Calos, 2004). Intλ's recombination mechanism takes the form of a tetrad complex of four Intλ monomers; each monomer binding to the DNA substrate at the same time (Rajeev et al, 2009). The reaction is started by the Intλ N-terminal domain of each monomer bound to the P and P' in *attPλ* (Biswas et al, 2005). The Intλ C-terminal domains (core binding and catalytic domains) recognize and bind to the complementary crossover sequences (COC' in the *attP* and BOB' in the *attB*) regions in a square planer formation, promoting DNA recombination (Van Duyne, 2005). One strand is cut by the first two active Intλ monomers at both the *attB and attP* sites. The next step involves DNA strand exchange and a ligation reaction between one strand from each site, reforming the integrity of the DNA strand. This strand exchange generates a Holliday junction (HJ) intermediate (Biswas et al, 2005). The HJ is resolved when the second pair of integrase monomers carry out the second DNA cut followed by the second DNA strand exchange and ligation, completing the integration event (Hsu & Landy, 1984). This recombination description is based on the "branch migration" model (Weisberg et al, 1983). DNA recombination results in the creation of two new, unique, hybrid sites, *attLλ* and *attRλ*, differing in sequence from either *attB* or *attP* (Stark et al, 1992). Once this reaction has occurred, expression of the site-specific recombinase is silenced by the action of λ repressor (CI) (Calendar, 2006; Serra-Moreno et al, 2008).

The lambdoid bacteriophage Φ24B is a Shiga toxin-encoding, temperate phage currently classified in the species *Escherichia virus 24B*, the genus *Traversvirus*, the subfamily *Sepvirinae* and the family *Podoviridae*. It has been well characterized with respect to host range (James et al, 2001), host receptor (Smith et al, 2007a), survival studies (Johannessen et al, 2005), comparative genetic composition (Smith et al, 2012; Smith et al, 2007b), induction signals (Bloch et al, 2015; Fogg et al, 2012; James et al, 2001; Licznerska et al, 2016), and DNA replication proficiency (Kozłowska et al, 2017; Nejman et al, 2011). The integrase of Φ24B and its cognate recombination directionality factor (RDF) have been identified and their expression characterized (Fogg et al, 2007; Fogg et al, 2011; Fogg et al, 2012). The Φ24B *int* gene (EF397940.1) encodes a product (IntΦ24B) of ~45 kDa whose expression is not controlled by the phage repressor, but is expressed constitutively (Fogg et al, 2011). This integrase, IntΦ24B, belongs to the tyrosine recombinase family (Balding et al, 2005; Esposito & Scocca, 1997; Fogg et al, 2007). Though various integration sites for the IntΦ24B have been identified (Fogg et al, 2007), and it is understood how prophage induction can be controlled with a constitutively expressed integrase and an inducible RDF (Fogg et al, 2011), little is known about the exact sequences and accessory factors outside the integrase and core crossover sites (Fogg et al, 2007) that are necessary to drive integration. In this study, we establish working protocols for both *in vitro* and *in situ* integration assays to determine the minimal required sequence lengths of both *attP* and *attB* to support integration at the primary integration site in the *E. coli* chromosome. We also investigate the role of the bacterial-encoded integration host factor (IHF) in prophage integration.

3 Results

Setting up an *in vitro* IntΦ24B assay

Φ24B is a member of the lambdoid group of phages based on its genome sequence and genomic context (Smith et al, 2012; Smith et al, 2007b). It was therefore reasonable to assume that IntΦ24B

would function in a similar fashion to Int$_{\lambda}$. To test this hypothesis, an *in vitro* assay was designed. Such an assay required at least 4 components: purified IHF, purified Int$_{\Phi 24B}$ and the DNA sequences *attP*$_{\Phi 24B}$, and *attB*$_{\Phi 24B}$. Conditions for the induction and purification of recombinant IHF have been previously published (Frumerie et al, 2005). However, the initial attempts to purify the recombinant IHF did not result in the purification of both subunits of the heterodimer. Therefore, the expression plasmid pEE2003 (Frumerie et al, 2005) was recovered and transformed into a different *E. coli* background, strain BL21, which enabled the purification of both IHF subunits (Fig.1A). Purified recombinant Int$_{\Phi 24B}$ was obtained from the previously described construct pΦ24$_B$-int (Fogg et al, 2011), which encodes an active, recombinant integrase with a histidine tag, inducible with arabinose (Fig.1B). Though the exact lengths of the *attB* and *attP* sequences needed to support recombination were unknown, it had been demonstrated that only *attB* and *attP* sites of 600 bp were sufficient to enable recombination *in situ* (Fogg et al, 2011). Therefore, different lengths of both *attP* and *attB* sequences were produced (Supplementary Tables 1 & 2) and cloned into pCR2.1 (Supplementary Table 1). The identity of transformants carrying potential *attB* constructs of 100, 200, 300, 400 or 600 bp in length (Fogg et al, 2007) or similarly designed *attP* constructs (100, 200, 300, 400 or 600 bp) were confirmed by PCR using M13 F/R primers and the introduction of potential point mutations ruled out by Sanger sequencing.

Before mixing the recombinant proteins with the DNA constructs, the absence of nucleases in the recombinant protein preparations was confirmed by incubating either IHF or Int$_{\Phi 24B}$ with 100 ng of pCR-Φ24$_B$-attP$_{600}$ (Fig. 2A). The *in vitro* integration assays were then set up to determine if Φ24$_B$ integrase was able to promote recombination between *attP* and *attB*. This assay was repeated multiple times with various reagent mixtures using different buffers, with or without the addition of crude cell lysate from *E. coli* TOP10 cells. Recombination events during the *in vitro* assays occurred only under very specific conditions: linearized pCR2.1-Φ24$_B$-attB$_{600}$, and supercoiled pCR2.1-Φ24$_B$-attP$_{600}$ in the presence of Int $_{\Phi 24B}$ and a crude cell extract of *E. coli* strain TOP10 (Fig. 2B & C). The identity of the recombination product was confirmed by PCR using M13 F/R or attRL F/R (attP150F/attBR) oligonucleotide primers (Supplementary Table 2) (Fig. 2D), and their identity was verified by Sanger sequencing. It was not clear from these results if the recombinant IHF protein was active or not or whether the assay conditions were not optimized to support recombination activity. Consequently, an *in situ* assay already demonstrated to work reliably (Fogg et al, 2011) was adapted to better characterize the requirements of Int$_{\Phi 24B}$.

Characterization of Int$_{\Phi 24B}$ requirements using an *in situ* assay

The *in situ* assay required that the cell hosting the recombination events was capable of stable maintenance of 3 different plasmids (Fig. 3A). The plasmid pΦ24$_B$-int replicates with an origin of replication from pBR322, which was also the same origin of replication supporting replication of the plasmids carrying *attB* and *attP* sites (Supplementary Table 1) used in the *in vitro* recombination assay. To avoid plasmid incompatibility issues in a single *E. coli* cell, all *attB* sequence variants were subcloned into pACYCDuet™-1 (Fig. 3C), which carries a p15A origin of replication. All *attP* sequence variants were subcloned into pCDFDuet™-1 (Fig. 3B), which carries an origin of replication unrelated to the other two plasmids (Supplementary Table 1). The appropriate plasmids were transformed into *E. coli* TOP10 cells (Supplementary Table 1), Int $_{\Phi 24B}$ expression was induced with arabinose, and integration events were identified by determining the existence of the large hybrid pΦ24B-*attRL* plasmid (Fig. 3A). The integration products were all confirmed by Sanger sequencing. The experiments were run with two replicates. The minimum *attP* sequence necessary to

support phage genome integration inside *E. coli* genome comprised between 350 bp and 427 bp, with almost equal integrase binding sites (P and P') site lengths (Fig. 3B).

The sequences required by Int$_{\Phi 24B}$ for activity with regards to *attB* were studied in a similar fashion. The large hybrid pΦ24B -*attRL* plasmid products (Fig. 3A) were identified indicating *attP* and *attB* recombination. The minimal *attB* sequence that could support phage integration comprised between 93 bp and 62bp (Fig 3C). The product identities were confirmed by Sanger sequencing. The *attB* site utilized by Int$_{\Phi 24B}$ is much longer than the 21 bp *attB*$_{\lambda}$ site utilized by Int$_{\lambda}$ (Mizuuchi & Mizuuchi, 1985).

The role of IHF in IntΦ24B activity

To test the importance of IHF in promoting the activity of Int $_{\Phi 24B}$ to drive integration of its *attP* into the *attB* in the *E. coli* genome, the *E. coli* strain TOP10 used in the *in situ* recombination assay was replaced with the *E. coli* strain JW1702-1 (Baba et al, 2006) in which the *himA* gene has been deleted. A deletion of one of the genes involved in the production of the IHF heterodimer ablates the production of an active IHF. The introduction of p$\Phi 24_B$-attB$_{288\text{-}288}$, p$\Phi 24_B$-attP$_{600}$ and p$\Phi 24_B$-int (Supplementary Table 1) into strain JW1702 resulted in the production of a large chimeric DNA product only after the expression of the Int$_{\Phi 24B}$ was induced by the addition of arabinose (data not shown). This demonstrated that integration could occur in the absence of IHF. However, the requirement of the crude cell lysate in the *in vitro* recombination assay does suggest that another, as of yet unidentified, host-encoded protein is required for Int $_{\Phi 24B}$ activity.

4 Discussion

Escherichia phage λ is generally used as a textbook illustration of temperate phage behavior. Our research shows that temperate phage integration, including the presence of host factors and recognition sites, is not generalizable from λ, in particular for the Shiga-toxin encoding phage $\Phi 24_B$. This study sought to better understand how $\Phi 24_B$ mediates integration and to define the factors required for Int$_{\Phi 24B}$ to function. We know that this integrase exhibits some degree of promiscuity, being able to utilize multiple different sequences in the *E. coli* genome (Fogg et al, 2007), so understanding how this enzyme functions and what factors are necessary for its function could lead to potential biotechnological applications. This knowledge also has important implications for understanding the dissemination of Stx phages and their ability to expand the diversity of Shiga toxin producing pathogens (Allison, 2007; Grotiuz et al, 2006; Muniesa et al, 2012).

Previous studies have demonstrated that experimental parameters required for various integrases in *in vitro* recombination assays are not uniform. In the case of recombination catalyzed by the bacteriophage P2-encoded integrase, a tyrosine recombinase, integration requires 2 *att* sequences, the phage integrase and a bacterial IHF (Sylwan et al, 2010), a phenomenon not uncommon across tyrosine recombinases (Groth & Calos, 2004). However, there are tyrosine recombinases like Cre, encoded by bacteriophage P1, that function without accessory proteins (Groth & Calos, 2004). Due to our lack of controls for the activity of recombinant IHF and Int$_{\Phi 24B}$, the only interpretable information obtained was that *in vitro*, recombination worked best if *attB* was linearized and *attP* was not (Fig. 2) and that proteins in a crude cell lysate were essential for recombination. Utilization of a linearized target for phage integration has been described previously, for the integrase encoded by the bacteriophage P2 (Frumerie et al, 2008), however Int$_{P2}$ and Int$_{\Phi 24B}$ have unrelated integrase box I and box II sites (Balding et al, 2005; Smith et al, 2007b). To better characterize the requirements of

Int$_{\Phi 24B}$, further experiments were run using a proven *in situ* assay (Fogg et al, 2011). The *in situ* assay does not allow the parameters of the recombination events to be strictly controlled in terms of the concentrations and ratios of DNA, protein, enzymes and reaction buffer content, but has been used to reliably detect recombination events (Bliska & Cozzarelli, 1987; Fogg et al, 2011; McCulloch et al, 1994; Rajeev et al, 2009).

Altering our approach to characterize Int$_{\Phi 24B}$ using an *in situ* assay enabled us to determine the minimal length of the *attP* sequence capable of supporting integration. The P' side of *attP*$_{\lambda}$ is ~80 bp in length, while the P side possesses the integrase binding site and is larger, ~140 bp (Goodman et al, 1999; Groth & Calos, 2004; Sylwan et al, 2010). A 600 bp DNA fragment harboring the *attP*$_{\Phi 24B}$ in its center had been shown previously (Fogg et al, 2011) to support integration, so this fragment was used and subsequently shortened to examine the impact of shorter sequence lengths on recombination (Fig. 4). In order to examine whether the P and P' arms flanking the central recombination core need to be equal sizes or not, *in situ* assays were run using pΦ24$_B$-attP$_{427}$, pΦ24$_B$-attP$_{350}$, pΦ24$_B$-attP$_{237}$ and pΦ24$_B$-attP$_{140}$ (Fig 4). The smallest *attP*$_{\Phi 24B}$ construct supporting recombination was pΦ24$_B$-attP$_{427}$. This is considerably larger than the *attP*$_{\lambda}$ site with *attP*$_{\Phi 24B}$ P and P' arm lengths more equal in size.

The *in situ* recombination assay also enabled determination of the minimal *attB* sequence requirements. Tyrosine site-specific recombinases have complex binding sites, however the *attB* site is usually less complex than the *attP* site. The site *attB*$_{\lambda}$, for example, is only 21 bp in length, 7 bp serving as the central crossover core sequence that interacts with the *attP* site (Mizuuchi et al, 1981) and 7 bp on either side serving as the B and B' arms. Nevertheless, according to our results, *attB*$_{\Phi 24B}$ site composition is not that simple. The shortest sequence tested that could support recombination was 93 bp long (Fig. 4), comprising of a 49 bp B arm followed, the 24 bp crossover site and a 20 bp long B' arm. This sequence is very long compared to other characterized tyrosine recombinase *attB* sites (Kolot et al, 1999; Lorbach et al, 2000). Some of the larger *attB* sequences characterized for members of the tyrosine recombinase family, *e.g.* FLP and Cre, require an *attB* sequence ~50 bp (Abremski & Hoess, 1984; Andrews et al, 1985), while the 62 bp *attB*$_{\Phi 24B}$ comprising of an 18 bp- B arm and a 20 bp B'arm could not support recombination. The smallest *attB*$_{\Phi 24B}$ site supporting recombination of an asymmetrical site. This would be very unusual for an *attB* site as most B and B' arms are the same size (Kolot et al, 1999; Lorbach et al, 2000), but the resolution of our strategy did not allow us to definitively determine whether the B arm is actually larger than the B' arm, though it is definitely unusual with regard to its length.

While bacterial-encoded IHF is necessary for the function of most tyrosine site-specific recombinase proteins (Fogg et al, 2014), there are integrases that do not require IHF, *e.g.* the integrase from the bacteriophage P22 of *Salmonella enterica* serovar Typhimurium can drive recombination in the absence of IHF protein, though IHF can enhance the binding of Int$_{P22}$ to *attP*$_{P22}$ (Smith-Mungo et al, 1994). Data from two different assays have demonstrated that Int$_{\Phi 24B}$ functions in the absence of IHF. Firstly, when the assay was performed *in situ* using the JW1702-1 strain of *E. coli* that cannot produce IHF due to the interruption of *himA* (encodes the IhfA subunit), recombination events were detected in an indistinguishable manner from those occurring in the TOP10 *E. coli* background. In line with data from the *in vitro* recombination assay, it became clear that recombinant IHF activity was not required, but that some other unknown factor(s) present in the *E. coli* crude protein extract were important. It is highly probable that the reason the *in vitro* assay did not work well under the conditions tested here was not because of the failure to produce biologically active recombinant proteins, but was due to the absence of at least one other currently undefined cellular factor/component.

The ability of Int$_{\Phi 24B}$ to drive phage genome integration inside the *E. coli* cell can have a profound impact on bacterial virulence (Allison, 2007; Fogg et al, 2012). The Stx phage integration results in a cell that, upon clonal expansion, generates numerous progeny that can produce Shiga toxin, the primary virulence factor of EHEC (Paton & Paton, 1998) which are a continuing threat to food safety and public health worldwide. Therefore, it is important to understand how this integrase functions and to identify precisely what is required for it to drive DNA recombination. Each Stx phage characterized to date carries a single integrase gene, and though there are many variants of integrase associated with Stx phages (Smith et al, 2012), Int$_{\Phi 24B}$ is particularly promiscuous being able to integrate efficiently into more than one location in the *E. coli* genome (Fogg et al, 2007). Thus far, only the primary *attB*$_{\Phi 24B}$ has been characterized; upon infection, $\Phi 24_B$ attempts to use this site first and will only use a secondary site if the primary site is occupied (Fogg et al, 2007). Comparison here (data not shown) of the now better defined primary *attB*$_{\Phi 24B}$ with the other sequences surrounding the secondary *attB* sites (Fogg et al, 2007) in the *E. coli* chromosome has not revealed any shared homologies beyond those identified by Fogg *et al.* (2007). More work is needed to identify the *E. coli*-encoded factors used by Int$_{\Phi 24B}$, how the integrase interacts with the secondary *attB* sites and what sequences within *attP* and *attB* actually support integrase binding. The work presented here builds upon our knowledge of the DNA sequences required for integration. We have demonstrated that the *attB* sequence needs to be linearized for integration to begin, and we have demonstrated that IHF is not required for Int$_{\Phi 24B}$ mediated recombination.

5 Materials and Methods

Strains, plasmids, media, buffers and oligonucleotides

All bacterial strains and plasmid constructs used in this study are described in appropriate detail in Supplementary Table 1. All oligonucleotide primers used in this study are listed in Supplementary Table 2.

Protein Overexpression

In order to produce recombinant IHF for the *in vitro* recombination assay, an overnight culture (1 mL) of *E. coli* TOP10 cells (Supplementary Table 1) harboring one of the two the histidine-tagged expression constructs, pEE2003 (kindly donated by E. Haggard-Ljungquist) or p$\Phi 24_B$-int, or BL21 cells harbouring pEE2003 was used to inoculate 10 mL of LB broth (3.7% (w/v) Luria Bertani Agar (Merck KGaA, Darmstadt, Germany) containing 100 μg/mL ampicillin that was incubated at 37℃ with shaking at 200 rpm until the OD_{600} reached 0.5-0.7. A sample (1 mL) was immediately harvested from the culture, and the cells recovered by centrifugation at 10,000 x g for 20-30 min, at 4℃ and resuspended in 50 μl of Tricine Sample Buffer (0.1 M Tris-Cl, pH 6.8; 24% (v/v) glycerol; 8% (w/v) SDS, 3.1% (w/v) DTT; 0.02% (w/v) Coomassie Blue R250). The resuspended cells were frozen at -20℃ for later use (non-induced control cells). Expression of the recombinant histidine tagged protein was induced from the remaining 9 mL of culture using either 0.02% w/v L-arabinose or 5 mM IPTG for recombinant histidine tagged Int$_{\Phi 24B}$ or IHF, respectively. The induced culture was further incubated at 37℃ with shaking at 200 r.p.m. for 3-4 hrs. The culture was then placed on ice and a sample (1 mL) was taken. The cells in that 1 mL were harvested by centrifugation at 10,000 x g for 20-30 min at 4℃, resuspended in 100 μl of Tricine Sample Buffer, and frozen at -20℃ until

needed (induced control). The rest of cells were harvested by centrifugation at 10,000 x g for 20-30 min at 4℃. The resultant pellet was resuspended carefully in 5 ml of ice-cold lysis buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, pH 8) for purification of the His-tagged protein under native conditions, according the instructions of QIAGEN's QIAexpressionist handbook (W. Sussex, U.K.). Using a MSE ultrasonic disintegrator (Henderson Biomedical, UK) with a microtip probe, the sample was treated 6-8 times with 10 seconds pulses at 200-300 Watts and an amplitude of 70 μm. The lysate was kept on ice at all times, until it was subjected to centrifugation at 10,000 x g for 20-30 min at 4°C. The supernatant was harvested (crude cell extract A, soluble proteins) and stored on ice. The resultant pellet was resuspended in 5 ml lysis buffer (crude extract B, insoluble proteins). Crude extracts A and B (5 µl, each), were mixed with 5 µl of Tricine Sample Buffer (62.5 mM Tris-HCl, pH 6.8; 2.5 % SDS; 0.002 % Bromophenol Blue; 0.7135 M (5%) β-mercaptoethanol; 10 % glycerol). These samples, along with the noninduced and induced control samples, were heated to 95℃ for 5 min. The samples were then centrifuged at 15,000 x g for 1 min, before being subjected to SDS-PAGE.

Protein purification

Histidine-tagged proteins were purified using the HisTrap (GE healthcare life sciences, Chicago, USA) protein purification 1ml column utilizing Ni-NTA technology. The column was first washed with water (5 x void volume), before equilibration with an equal volume of lysis buffer. The column was loaded with crude extract A (soluble proteins and then washed with several washing buffers (50 mM NaH_2PO_4, 300 mM NaCl, pH 8) comprised of varying imidazole concentrations (20, 30, 40, 50 and 60 mM), and finally treated with elution buffer (50 mM NaH_2PO_4, 300 mM NaCl, pH 8 with 250 mM imidazole). The volume of washing buffers and elution buffer used were 5 times the column void volume. The washes and eluates were collected in graded test tubes. The column was then washed with lysis buffer, water and finally filled with 20% ethanol for storage at 4 ℃ or subsequent reuse. Protein samples were dialyzed to remove all imidazole using dialysis tubing (Medicell, London, UK) with a pore-size capable of retaining proteins of ≥12-14 kDa and then concentrated using Vivaspin 20 tubes (5 kDa) (GE Healthcare, Buckinghamshire, UK). The concentration of each protein was measured using the Pierce BCA Protein Assay Kit according to the manufacturer's instructions. A sample of 15 µl from each elution tube was mixed with 5ul of Tricine Sample Buffer, heated to 95℃ for 5 min and loaded onto an SDS-PAGE gel.

SDS PAGE

Sodium dodecyl sulphate polyacrylamide gel electrophoresis, SDS-PAGE (Laemmli, 1970), was performed to examine protein preparations using the Mini-Protean 3 Unit (Bio Rad, Serial NO 67S). The vertical running gel was 0.75 mm thick and composed of a bottom Separating gel (1 ml water; 1.25 ml 3M Tris-HCl/SDS, pH 8.4; 1.12 ml 40% acrylamide; 0.37 ml glycerol; 5 μl 30% APS; and 5 μl TEMED), and an upper Stacking gel (2.03 ml water; 0.77 ml 3M Tris-HCl/SDS, pH 8.4; 0.31 ml 40% acrylamide; 5 μl 30% APS; and 5 μl TEMED). A 20 μl-sized well forming comb was used to cover the top of the stacking gel until it was polymerized. The polymerized gel was placed into the electrophoresis tank, with appropriate cathode (0.5 M Tris base, pH8.45) and anode (1 M Tricine, pH 8.3 with 2% SDS) buffers. Samples (15 μl sample and 5 μl Tricine sample buffer) were loaded in the gel wells using a pre-stained 10-250 kD molecular weight marker (BioRad) in the outer wells. electrophoresis was performed at 150 V for 30 min. The unit was dismantled, and the gel placed in Coomassie Blue Stain (2% (w/v) Coomassie Blue R250 in 45% (v/v) methanol and 10% (v/v/) glacial acetic acid) with gentle agitation for 30 min. The gel was destained with several exchanges of

Coomassie Blue Destain (45% (v/v) methanol and 10% (v/v/) glacial acetic acid) until a clear gel background was obtained.

Cloning of bacterial and phage attachment sites in pCR2.1

The *attB* core sequence (Fogg et al, 2007) and either 100 bp, 200 bp, 300 bp, 400 bp or 600 bp of its flanking sequence were amplified from the MC1061 *E coli* genome using *attB*100 F/R, *attB* 200 F/R, *attB*300 F/R *attB*400 F/R or attB 600F/R oligonucleotide primers, respectively.
At the same time, the *attP* core sequence (Fogg et al, 2007) and either 100 bp, 200 bp, 300 bp, 400 bp or 600 pb of flanking sequence was amplified from the $\Phi 24_B$ genome using *attP*140 F/R, *attP*200 F/R *attP*300 F/R *attP*400 F/R or *attP* 600F/R primers, respectively. Amplification parameters involved an initial denaturation period of 2 min at 95°C followed by 30 cycles of denaturation for 10 sec. at 95°C, annealing for 10 sec. at 50°C, extension for 30 sec. at 72°C, followed by a final extension period for 2 min. at 72°C. Amplification reactions used MyTaq™ (Bioline) DNA polymerase according to the manufacturer's recommendations. Each of the *attB* or *attP* sequences was cloned, separately, into the pCR2.1 cloning vector (Invitrogen), according to manufacturer's guidelines. The *attB* and *attP* pCR2.1 constructs, were all individually transformed into chemically competent TOP10 cells. The transformants were cultured overnight on LB agar plates containing ampicillin (100 μg/mL) and kanamycin (50 μg/mL). The identities of the desired transformants were confirmed by colony PCR using primers (*M13 F/R*) flanking the multiple cloning site of the pCR2.1 with the amplification parameters described earlier before they were sent for Sanger sequencing. Freezer stocks were made from each positive transformant and stored at -80 °C.

In Vitro Recombination Assay

In order to prepare a crude cellular protein extract, cells from an overnight TOP10 culture were harvested by centrifugation for 20-30 min at 10,000 x g at 4℃. The resultant pellet was resuspended carefully in 5 ml of ice-cold lysis buffer. The cells were lysed as described above. The lysate was kept on ice at all times, until it was subjected to centrifugation for 20-30 min at 10,000g at 4°C. Total protein in the supernatant was quantified as described above and used as a crude protein extract.

Recombination assays were always performed in duplicate using either coiled p*attP* plasmid (pCR2.1-$\Phi 24_B$-attP$_{100, 200, 300, 400 \text{ or } 600,}$ Supplementary Table 1) with coiled p*attB* plasmid (pCR2.1-$\Phi 24_B$-attB$_{100. 200, 300, 400 \text{ or } 600,}$ Supplementary Table 1), coiled p*attP* plasmid with linear p*attB* plasmid or linear p*attP* plasmid with coiled p*attB* plasmid. Approximately 500 ng of each DNA target species were incubated with or without combinations of recombinant IHF (100 ng ~90% purity as determined by SDS-PAGE), recombinant Int$_{\Phi 24_B}$ (200 ng 90% purity as determined by SDS-PAGE) and/or *E. coli* derived crude protein extracts (500 ng protein) in a final reaction volume of 50 μl. pCR2.1 plasmids harboring p*attP* or p*attB* sequences were linearized by the restriction enzyme *Bam*H I. Two types of buffer were utilized in these reactions: Buffer 1 (20 mM Tris, pH 7.5, 100 mM NaCl, 1% glycerol, 0.1 mM EDTA (Thorpe & Smith, 1998)) or Buffer 2 (50 mM Tris-Cl, pH 7.8, 60 mM KCl, 250 µg of BSA, 0.5 mM EDTA, 10% glycerol 5 mM spermidine in a final reaction volume of 15 µl (Goodman et al, 1999)). The reactions were incubated for 6 or 24 hours at 37◦C and stopped by separating the resulting DNA molecules by TAE agarose gel electrophoresis. Any band of the expected recombinant size was excised from the gel, extracted using an Isolate PCR and Gel kit from Bioline, and used as PCR template. If recombination occurred, a unique *attRL* product could be amplified using attRL F/R primers and the amplification parameters described above. The identity of the amplified product was confirmed by Sanger sequencing (GATC Biotech AG [Germany]).

Subcloning *attB* and *attP* sites into compatible plasmids

The plasmid construct, pCR-Φ24B-*attP*$_{600}$ harboring the *attP* core and 600 bp of flanking sequence, was purified from *E. coli* TOP10 strain. The *attP* fragment was recovered using *Bam*H I and *Xho* I restriction endonuclease enzymes. The same pair of enzymes was used to cut the high copy number plasmid, pCDF-Duet. After endonuclease inactivation, by incubation at 65 °C for 20 minutes, the digestion products were each purified on a TAE agarose gel and recovered using the Isolate PCR and Gel kit (Bioline, London, U.K.). The *attP* sequence was ligated into pCDF-Duet to form pΦ24$_B$-*attP*$_{600}$), before it was transformed into competent TOP10 *E. coli* cells. In parallel pCR-Φ24$_B$-attB$_{600}$ harboring the *attB* core and 600 bp of flanking sequence, and the medium copy number pACYC-Duet, a plasmid compatible with pCDF-Duet, were also digested using the same endonuclease enzymes (*Bam*H I and *Xho* I) and the products cloned together as described above to produce pΦ24$_B$-attB$_{600}$ before it was transformed into fresh TOP10 competent cells. This cloning strategy was used repeatedly to produced constructs harboring variously sized *attB* and *attP* sequences.

In situ recombination assay

All *in situ* recombination assays were performed inside One Shot TOP10 *E coli* cells (Supplementary Table 1). Each required plasmid transformed sequentially into these cells until all required, compatible plasmids for each assay were in one *E coli* host cell. An appropriate amount of LB broth, supplemented with the plasmid's resistance antibiotics at their specific concentrations, was inoculated with 1 ml of an overnight culture of cells harboring the compatible plasmids. This culture was incubated at 37°C with shaking at 200 rpm until the OD$_{600}$ 0.5-0.8 was reached. Expression of the integrase protein (Int) was induced by the addition of L-Arabinose to a final concentration of 0.02% w/v. The cultures were incubated for an additional three hours at same incubation temperature. Finally, from each assay flask, 5 ml samples were taken and all plasmid DNA present was extracted from the cells in the samples. Recombination events during the *in situ* recombination assays were only determined to have occurred when the following three criteria were met: 1) a PCR reaction using *att*RL F/R primers resulted in the production of DNA product, 2) TAE agarose gel electrophoresis verified the size of the amplification product and 3) Sanger sequencing of the PCR products confirmed the identity of the hybrid site. All these assays where performed along with a negative control comprised of a sample that was not induced with arabinose so that no Int was produced.

6 Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

7 Author Contributions

HEA and AJM conceived the project, oversaw its execution and contributed to problem solving. MRM performed all of the experiments described in the manuscript and made the unique constructs described therein. EMA assisted in experimental design and execution. All authors were involved in writing the manuscript.

8 Funding

HEA and AJM gratefully acknowledge the support of the Biotechnology and Biological Sciences Research Council (BBSRC) BB/I013431/1. E.M.A gratefully acknowledges the support of the Biotechnology and Biological Sciences Research Council (BBSRC); this research was funded in part by the BBSRC Institute Strategic Programme Gut Microbes and Health BB/R012490/1 and its constituent project BBS/E/F/000PR10356. MRM acknowledges the support of both the Higher Committee for Education Development in Iraq (HCED) and the University of Anbar-Iraq for the funding he has received that enabled this project.

9 Acknowledgments

We are grateful to members of LabH for helpful discussions, especially to Paul Loughnane for his expert technical assistance. We acknowledge the assistance of Paul Fogg for providing us with the plasmid pBAD/Myc-His C, possessing the Φ24B *int* gene. Finally, we are grateful to E. Haggard-Ljungquist for providing us with the plasmid pEE2003 that carries the *himA* and *himD* genes of the *E. coli* IHF. The content of this manuscript has been published in part as part of the thesis of Mohammed Rahdi Mohaisen (Mohaisen, 2017).

10 Supplementary Material

Supplementary Table 1. Bacterial strains and plasmids. and Supplementary Table 2. Oligonucleotide primers used for PCR amplification of DNA.

11 References

Abremski, K. & Hoess, R. (1984) Bacteriophage P1 site-specific recombination. Purification and properties of the Cre recombinase protein. *J Biol Chem*, 259(3), 1509-14.

Allison, H. E. (2007) Stx-phages: drivers and mediators of the evolution of STEC and STEC-like pathogens. *Future Microbiol*, 2(2), 165-74.

Andrews, B. J., Proteau, G. A., Beatty, L. G. & Sadowski, P. D. (1985) The FLP recombinase of the 2 micron circle DNA of yeast: interaction with its target sequences. *Cell*, 40(4), 795-803.

Azam, T. A. & Ishihama, A. (1999) Twelve species of the nucleoid-associated protein from *Escherichia coli*. Sequence recognition specificity and DNA binding affinity. *J Biol Chem*, 274(46), 33105-13.

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K. A., Tomita, M., Wanner, B. L. & Mori, H. (2006) Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol*, 2, 2006.0008.

Balding, C., Bromley, S. A., Pickup, R. W. & Saunders, J. R. (2005) Diversity of phage integrases in *Enterobacteriaceae*: development of markers for environmental analysis of temperate phages. *Environ Microbiol*, 7(10), 1558-67.

Biswas, T., Aihara, H., Radman-Livaja, M., Filman, D., Landy, A. & Ellenberger, T. (2005) A structural basis for allosteric control of DNA recombination by lambda integrase. *Nature*, 435(7045), 1059-66.

Bliska, J. B. & Cozzarelli, N. R. (1987) Use of site-specific recombination as a probe of DNA structure and metabolism in vivo. *J Mol Biol*, 194(2), 205-18.

475 Bloch, S., Nejman-Faleńczyk, B., Topka, G., Dydecka, A., Licznerska, K., Narajczyk, M., Necel, A.,
476 Węgrzyn, A. & Węgrzyn, G. (2015) UV-Sensitivity of Shiga Toxin-Converting Bacteriophage
477 Virions Φ24B, 933W, P22, P27 and P32. *Toxins (Basel)*, 7(9), 3727-39.

478 Calendar, R. (2006) *The bacteriophages*, 2nd ed. edition. New York ; Oxford: Oxford University
479 Press.

480 Esposito, D. & Scocca, J. J. (1997) The integrase family of tyrosine recombinases: evolution of a
481 conserved active site domain. *Nucleic Acids Res*, 25(18), 3605-14.

482 Fogg, P. C., Colloms, S., Rosser, S., Stark, M. & Smith, M. C. (2014) New applications for phage
483 integrases. *J Mol Biol*, 426(15), 2703-16.

484 Fogg, P. C., Gossage, S. M., Smith, D. L., Saunders, J. R., McCarthy, A. J. & Allison, H. E. (2007)
485 Identification of multiple integration sites for Stx-phage ϕ24B in the *Escherichia coli* genome,
486 description of a novel integrase and evidence for a functional anti-repressor. *Microbiol.*, 153(Pt 12),
487 4098-4110.

488 Fogg, P. C., Rigden, D. J., Saunders, J. R., McCarthy, A. J. & Allison, H. E. (2011) Characterization
489 of the relationship between integrase, excisionase and antirepressor activities associated with a
490 superinfecting Shiga toxin encoding bacteriophage. *Nucleic Acids Res*, 39(6), 2116-29.

491 Fogg, P. C., Saunders, J. R., McCarthy, A. J. & Allison, H. E. (2012) Cumulative effect of prophage
492 burden on Shiga toxin production in *Escherichia coli*. *Microbiology*, 158(Pt 2), 488-97.

493 Frumerie, C., Sylwan, L., Ahlgren-Berg, A. & Haggård-Ljungquist, E. (2005) Cooperative
494 interactions between bacteriophage P2 integrase and its accessory factors IHF and Cox. *Virology*,
495 332(1), 284-94.

496 Frumerie, C., Sylwan, L., Helleday, T., Yu, A. & Haggård-Ljungquist, E. (2008) Bacteriophage P2
497 integrase: another possible tool for site-specific recombination in eukaryotic cells. *J Appl Microbiol*,
498 105(1), 290-9.

499 Goodman, S. D., Velten, N. J., Gao, Q., Robinson, S. & Segall, A. M. (1999) In vitro selection of
500 integration host factor binding sites. *J Bacteriol*, 181(10), 3246-55.

501 Groth, A. C. & Calos, M. P. (2004) Phage integrases: biology and applications. *J Mol Biol*, 335(3),
502 667-78.

503 Grotiuz, G., Sirok, A., Gadea, P., Varela, G. & Schelotto, F. (2006) Shiga toxin 2-producing
504 *Acinetobacter haemolyticus* associated with a case of bloody diarrhea. *J Clin Microbiol*, 44(10),
505 3838-41.

506 Hsu, P. L. & Landy, A. (1984) Resolution of synthetic att-site Holliday structures by the integrase
507 protein of bacteriophage lambda. *Nature*, 311(5988), 721-6.

508 Hsu, P. L., Ross, W. & Landy, A. (1980) The lambda phage *att* site: functional limits and interaction
509 with Int protein. *Nature*, 285(5760), 85-91.

510 James, C. E., Stanley, K. N., Allison, H. E., Flint, H. J., Stewart, C. S., Sharp, R. J., Saunders, J. R. &
511 McCarthy, A. J. (2001) Lytic and lysogenic infection of diverse *Escherichia coli* and *Shigella* strains
512 with a verocytotoxigenic bacteriophage. *Appl Environ Microbiol*, 67(9), 4335-7.

513 Johannessen, G. S., James, C. E., Allison, H. E., Smith, D. L., Saunders, J. R. & McCarthy, A. J.
514 (2005) Survival of a Shiga toxin-encoding bacteriophage in a compost model. *FEMS Microbiol Lett*,
515 245(2), 369-75.

Kikuchi, Y. & Nash, H. A. (1978) The bacteriophage lambda int gene product. A filter assay for genetic recombination, purification of int, and specific binding to DNA. *J Biol Chem*, 253(20), 7149-57.

Kolot, M., Silberstein, N. & Yagil, E. (1999) Site-specific recombination in mammalian cells expressing the Int recombinase of bacteriophage HK022. *Mol Biol Rep*, 26(3), 207-13.

Kozłowska, K. I., Tymecka-Mulik, J. & Węgrzyn, G. (2017) Purified Stx and λ phage initiator O proteins bind specifically to two different origins of replication in vitro. *Protein Expr Purif*, 131, 16-26.

Laemmli, U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, 227(5259), 680-5.

Licznerska, K., Dydecka, A., Bloch, S., Topka, G., Nejman-Faleńczyk, B., Węgrzyn, A. & Węgrzyn, G. (2016) The Role of the *exo-xis* region in oxidative stress-mediated induction of Shiga toxin-converting prophages. *Oxid Med Cell Longev*, 2016, 8453135.

Lorbach, E., Christ, N., Schwikardi, M. & Droge, P. (2000) Site-specific recombination in human cells catalyzed by phage lambda integrase mutants. *J Mol Biol*, 296(5), 1175-81.

Madsen, P. L., Johansen, A. H., Hammer, K. & Brøndsted, L. (1999) The genetic switch regulating activity of early promoters of the temperate lactococcal bacteriophage TP901-1. *J Bacteriol*, 181(24), 7430-8.

McCulloch, R., Coggins, L. W., Colloms, S. D. & Sherratt, D. J. (1994) Xer-mediated site-specific recombination at *cer* generates Holliday junctions in vivo. *EMBO J*, 13(8), 1844-55.

Mizuuchi, K., Weisberg, R., Enquist, L., Mizuuchi, M., Buraczynska, M., Foeller, C., Hsu, P. L., Ross, W. & Landy, A. (1981) Structure and function of the phage lambda *att* site: size, Int-binding sites, and location of the crossover point. *Cold Spring Harb Symp Quant Biol*, 45 Pt 1, 429-37.

Mizuuchi, M. & Mizuuchi, K. (1980) Integrative recombination of bacteriophage lambda: extent of the DNA sequence involved in attachment site function. *Proc Natl Acad Sci U S A*, 77(6), 3220-4.

Mizuuchi, M. & Mizuuchi, K. (1985) The extent of DNA sequence required for a functional bacterial attachment site of phage lambda. *Nucleic Acids Res*, 13(4), 1193-208.

Mohaisen, M. R. (2017) *Molecular characterization of the activity and requirements of a novel and promiscuous bacteriophage integrase*. Ph.D. University of Liverpool, October 2017.

Muniesa, M., Hammerl, J. A., Hertwig, S., Appel, B. & Brüssow, H. (2012) Shiga toxin-producing *Escherichia coli* O104:H4: a new challenge for microbiology. *Appl Environ Microbiol*, 78(12), 4065-73.

Nejman, B., Nadratowska-Wesołowska, B., Szalewska-Pałasz, A., Węgrzyn, A. & Węgrzyn, G. (2011) Replication of plasmids derived from Shiga toxin-converting bacteriophages in starved *Escherichia coli*. *Microbiology*, 157(Pt 1), 220-33.

O'Brien, A. D., Newland, J. W., Miller, S. F., Holmes, R. K., Smith, H. W. & Formal, S. B. (1984) Shiga-like toxin-converting phages from *Escherichia coli* strains that cause hemorrhagic colitis or infantile diarrhea. *Science*, 226(4675), 694-6.

Paton, A. W. & Paton, J. C. (1998) Detection and characterization of Shiga toxigenic *Escherichia coli* by using multiplex PCR assays for *stx1*, *stx2*, *eaeA*, enterohemorrhagic *E. coli hlyA*, rfbO111, and rfbO157. *J Clin Microbiol*, 36(2), 598-602.

Rajeev, L., Malanowska, K. & Gardner, J. F. (2009) Challenging a paradigm: the role of DNA homology in tyrosine recombinase reactions. *Microbiol Mol Biol Rev*, 73(2), 300-9.

Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular cloning : a laboratory manual*, 2nd edition. Cold Spring Harbor, N.Y: Cold Spring Harbor Laboratory.

Sanyal, S. J., Yang, T. C. & Catalano, C. E. (2014) Integration host factor assembly at the cohesive end site of the bacteriophage lambda genome: implications for viral DNA packaging and bacterial gene regulation. *Biochemistry*, 53(48), 7459-70.

Saunders, J., Allison, H., James, C., McCarthy, A. & Sharp, R. (2001) Phage-mediated transfer of virulence genes. *J Chem Tech Biotech*, 76(7), 662-666.

Serra-Moreno, R., Jofre, J. & Muniesa, M. (2008) The CI repressors of Shiga toxin-converting prophages are involved in coinfection of *Escherichia coli* strains, which causes a down regulation in the production of Shiga toxin 2. *J Bacteriol*, 190(13), 4722-35.

Smith, D. L., James, C. E., Sergeant, M. J., Yaxian, Y., Saunders, J. R., McCarthy, A. J. & Allison, H. E. (2007a) Short-tailed Stx phages exploit the conserved YaeT protein to disseminate Shiga toxin genes among enterobacteria. *J Bacteriol*, 189(20), 7223-33.

Smith, D. L., Rooks, D. J., Fogg, P. C., Darby, A. C., Thomson, N. R., McCarthy, A. J. & Allison, H. E. (2012) Comparative genomics of Shiga toxin encoding bacteriophages. *BMC Genomics*, 13, 311.

Smith, D. L., Wareing, B. M., Fogg, P. C., Riley, L. M., Spencer, M., Cox, M. J., Saunders, J. R., McCarthy, A. J. & Allison, H. E. (2007b) Multilocus characterization scheme for Shiga toxin-encoding bacteriophages. *Appl Environ Microbiol*, 73(24), 8032-40.

Smith-Mungo, L., Chan, I. T. & Landy, A. (1994) Structure of the P22 att site. Conservation and divergence in the lambda motif of recombinogenic complexes. *J Biol Chem*, 269(32), 20798-805.

Stark, W. M., Boocock, M. R. & Sherratt, D. J. (1992) Catalysis by site-specific recombinases. *Trends Genet*, 8(12), 432-9.

Sugimura, S. & Crothers, D. M. (2006) Stepwise binding and bending of DNA by *Escherichia coli* integration host factor. *Proc Natl Acad Sci U S A*, 103(49), 18510-4.

Sylwan, L., Frumerie, C. & Haggård-Ljungquist, E. (2010) Identification of bases required for P2 integrase core binding and recombination. *Virology*, 404(2), 240-5.

Thorpe, H. M. & Smith, M. C. (1998) In vitro site-specific integration of bacteriophage DNA catalyzed by a recombinase of the resolvase/invertase family. *Proc Natl Acad Sci U S A*, 95(10), 5505-10.

Unkmeir, A. & Schmidt, H. (2000) Structural analysis of phage-borne *stx* genes and their flanking sequences in Shiga toxin-producing *Escherichia coli* and *Shigella dysenteriae* type 1 strains. *Infect Immun*, 68(9), 4856-64.

Van Duyne, G. D. (2005) Lambda integrase: armed for recombination. *Curr Biol*, 15(17), R658-60.

Wagner, P. L., Livny, J., Neely, M. N., Acheson, D. W., Friedman, D. I. & Waldor, M. K. (2002) Bacteriophage control of Shiga toxin 1 production and release by *Escherichia coli*. *Mol Microbiol*, 44(4), 957-70.

Weisberg, R. A., Enquist, L. W., Foeller, C. & Landy, A. (1983) Role for DNA homology in site-specific recombination. The isolation and characterization of a site affinity mutant of coliphage lambda. *J Mol Biol*, 170(2), 319-42.

598 Yang, S. W. & Nash, H. A. (1995) Comparison of protein binding to DNA *in vivo* and *in vitro*:
599 defining an effective intracellular target. *EMBO J*, 14(24), 6292-300.

600 Łobocka, M. B., Rose, D. J., Plunkett, G., Rusin, M., Samojedny, A., Lehnherr, H., Yarmolinsky, M.
601 B. & Blattner, F. R. (2004) Genome of bacteriophage P1. *J Bacteriol*, 186(21), 7032-68.

602

12 Figure Legends

Figure 1. Purification of recombinant proteins. Panel A, Production of dimeric IHF from the construct pEE2003. *E. coli* strain BL21 carrying the plasmid pEE203 (Supplementary Table 1) was cultured in 10 mL of LB broth and overexpression was induced with 5 mM IPTG. After 3-4 hours the cells were harvested, and the proteins purified using a HisTrap (1ml) column (GE Healthcare Life Sciences). The various steps of this process were analysed on an SDS-PAGE gel. Lanes: **MW**, prestained molecular weight markers (2-250 kD) (BioRad); **1**, whole cell lysate; **2**, Ni-column flow through; **3-5**, column washes; **6**, IHF in 60 mM imidazole washing buffer**.** Arrows indicate the 11.3 kDa his-tagged IhfA subunit and the co-purifying 10.6 kDa IhfB subunit. **Panel B, Production and Purification of recombinant Int Φ24B.** *E. coli* strain MC1061 carrying the plasmid pΦ24B-int (Supplementary Table 1) was cultured in 10 mL of LB broth and overexpression was induced with L-Arabinose. After 3-4 hours the cells were harvested, and the proteins purified using a HisTrap (1ml) column (GE Healthcare Life Sciences). The various steps of this process were analysed on an SDS-PAGE gel. Lanes: MW, pre-stained molecular weight markers (10-250 kDa; BioRad); 1, whole cell lysate; 2, Ni-column flow through; 3-5, column washes; 6, Integrase in 60 mM imidazole. The arrow indicates the purified 45 kDa monomer of IntΦ24B.

Figure 2. Data from the *in* vitro recombination assay. Panel A, Demonstration that the recombinant IHF and Integrase preparations did not contain nucleases. Purified recombinant IHF and IntΦ24B were incubated with pCR2.1- Φ24B-*attP*600 to monitor these protein preparations for possible co-purifying nucleases. Lanes: **MW**, HyperLadder™ 1 kb (Bioline); **1**, 100 ng pCR2.1-Φ24B-*attP*600; **2 and 3**, 100 ng pCR2.1- Φ24B-*attP*600 with 1 µg of IHF or 1 µg of IntΦ24B, respectively, incubated at 30 °C for 1 hr. No co-purifying nuclease activity was detected. **Panels B-C, *In vitro* recombination assay results using crude protein extracts. B,** results after incubation for 6 hr at 37°C. **C**, results after incubation for 24 hr at 37°C; Reaction composition by lane numbers: MW, HyperLadder™ 1 kb (Bioline)**1 & 4** , supercoiled pCR2.1- Φ24B-*attP*600 only, without crude protein extract **2 & 5**, linearized pCR2.1-Φ24B-*attB*600 only, without crude protein extract **3 & 6**, linearized pCR2.1-Φ24B-*attB*600 and supercoiled pCR2.1- Φ24B-*attP*600 with 500 ng crude protein extract from TOP10 *E. coli*; **7**, supercoiled pCR2.1- Φ24B-*attP*600 and linearized pCR2.1-Φ24B-*attB*600, only. Recombination was only detected in Lanes 3 and 6, by the presence of a 9 kb band. **Panel D, PCR assay of the recombination products. MW**, HyperLadder™ 1 kb (Bioline); **1**, positive control using M13 F/R primers with pCR2.1-*attP600* as template; **2**, *attRL* PCR fragment amplified using M13 F/R primers; **3**, *attRL* PCR fragment amplified using *attP*150 F/attB R primers**; 4.** negative control using M13 F/R primers without template.

Figure 3. Determining the minimal attP and attB sites utilized by IntΦ24B. Panel A. Diagram depicting the *in situ* assay set up, arabinose induction of IntΦ24B expression, and the resulting chimeric plasmid. **Panel B.** The design of the various attP constructs (plasmids with nomenclature p*Φ*24B-attP, Suppl Table 1), providing information on the lengths of the P (solid purple) and P' (hashed purple) sequences and drawn to scale as they flank the 24 bp central crossover site (green). The detection of recombination (+) or absence (-) is also indicated. **Panel C.** The design of the various attB constructs (plasmids with nomenclature p*Φ*24B-attB, Suppl Table 1), providing information on the lengths of the B (solid blue) and B' (hashed blue) sequences and drawn to scale as they flank the 24 bp central crossover site (red). The detection of recombination (+) or absence (-) is also indicated.

647
648
649